

Dose–response and threshold-mediated mechanisms in mutagenesis: statistical models and study design

David P. Lovell ^{a,1}

BIBRA International, Woodmansterne Road, Carshalton, Surrey SM5 4DS, UK

Received 30 November 1998; received in revised form 11 May 1999; accepted 8 June 1999

Abstract

The objective of this paper is to review the use, in mutagenesis, of various mathematical models to describe the dose–response relationship and to try to identify thresholds. It is often taken as axiomatic that genotoxic carcinogens could damage DNA at any level of exposure, leading to a mutation, and that this could ultimately result in tumour development. This has led to the assumption that for genotoxic chemicals, there is no discernible threshold. This assumption is increasingly being challenged in the case of aneugens. The distinction between ‘absolute’ and ‘pragmatic’ thresholds is made and the difficulties in determining ‘absolute’ thresholds using hypothesis testing approaches are described. The potential of approaches, based upon estimation rather than statistical significance for the characterization of dose–response relationships, is stressed. The achievement of a good fit of a mathematical model to experimental data is not proof that the mechanism supposedly underlying this model is operating. It has been argued, in the case of genotoxic chemicals, that any effects produced by a genotoxic chemical which augments that producing a background incidence in unexposed individuals will lead to a dose–response relationship that is non-thresholded and is linear at low doses. The assumptions underlying this presumption are explored in the context of the increasing knowledge of the mechanistic basis of mutagenicity and carcinogenicity. The possibility that exposure to low levels of genotoxic chemicals may induce and enhance defence and repair mechanisms is not easily incorporated into many of the existing mathematical models and should be an objective in the development of the next generation of biologically based dose–response (BB-DR) models. Studies aimed at detecting or characterizing non-linearities in the dose–response relationship need appropriate experimental designs with careful attention to the choice of biomarker, number and selection of dose levels, optimum allocation of experimental units and appropriate levels of replication within and repetition of experiments. The characterization of dose–response relationships with appropriate measures of uncertainty can help to identify ‘pragmatic’ thresholds based upon biologically relevant criteria which can help in the regulatory process. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Thresholds; Mutagenicity; Aneuploidy; Mathematical models; Hypothesis testing; Experimental design

1. Introduction

The objective of this paper is to review the use, in mutagenesis, of various mathematical models to de-

¹ Tel.: +44-1304-643554; fax: +44-1304-658339; e-mail: david_lovell@sandwich.pfizer.com

¹ Present address: Biometrics Department, Central Research, Pfizer Ltd., Ramsgate Road, Sandwich, Kent CT13 9NJ, UK.

scribe the dose–response relationship and to try to identify thresholds. It has usually been considered axiomatic that chemicals which damage the genetic material produce adverse effects which are non-thresholded. It is assumed that these chemicals can damage DNA at any level of exposure and such damage could result in tumour development. In this model of carcinogenesis, it is believed that there is a finite probability that a single molecule of a carcinogen may be sufficient to damage the genetic material irreversibly. Such a mutated cell may subsequently proliferate, eventually leading to a tumour. This model of no-safe-dose of a carcinogen or, alternatively, the single molecule theory of carcinogenesis is closely related to the one-hit models developed for investigating the low-dose effects of ionizing radiation [1]. The no-safe-dose of radiation has been a basic principle of radiation biology as developed by organisations such as the US National Council on Radiation Protection (NCRP) [2,3].

It is extremely difficult, if not impossible, to test the hypothesis of the no-threshold model of carcinogenesis because it has all the philosophical problems of trying, in effect, to design an experiment to prove a negative, i.e. that there is a threshold below which there are doses of a carcinogen where no adverse effects occur. Conventional rodent carcinogenicity bioassays lack the statistical power to resolve whether a threshold may exist in carcinogenesis while the results from larger-scale experiments deliberately designed to test for effects of carcinogens at low doses have proved to be equivocal. In many senses, the debate is unproductive and an example of what Weinberg [4] calls trans-science: the formulation of a problem in scientific terms without the ability to investigate the problem using an experimental approach. However, the increasing availability of a wide range of biomarkers provides the opportunity to define responses at dose levels closer to possible human exposures.

Thresholds, however, are an accepted principle in pharmacology for many receptor-activation-dependent processes such as nerve fibre firing. The argument for the existence of thresholds for toxicological processes is based upon homeostatic mechanisms which allow an organism to compensate for and adapt to toxicological insults. These mechanisms include the ability to detoxify a chemical before it

reaches a critical target and to repair damage from low-level exposure before adverse effects occur. True threshold dose levels can also be postulated which have to be exceeded before, e.g., cytotoxicity or the saturation of detoxification pathways occurs.

2. Mutagenesis and aneuploidy

Mutagenesis includes a range of genetic damage including gene mutation, chromosomal damage and changes in chromosomal number. Such damage may arise either from direct damage to DNA or indirectly by effects on other targets such as proteins. The axiom that, chemicals which cause genetic damage have no thresholds, is being challenged for aneugens (chemicals which induce aneuploidy). Aneugens may act by effects on non-DNA targets such as the inhibition of spindle function with no direct interaction of the aneugen with the DNA. It has been argued that these mechanisms have a threshold of action in contrast to DNA-reactive mutagens where covalent binding to DNA occurs and which are assumed to be non-thresholded.

A postulated mechanistic basis for a threshold for aneuploidy is by binding to a target receptor such as tubulin. It is argued that a critical number of target sites must be occupied before the biological effect occurs, such that there is a degree of redundancy in this target. This binding to the target receptor is also usually reversible, such that repeated exposure at low doses is not additive, resulting in a dose–response with a threshold [5].

Any carcinogenicity or mutagenicity resulting from aneuploidy is, thus, considered a secondary consequence of a toxicological event that has a threshold; this should then be incorporated into risk assessment procedures. Aardema et al. [6], for example, note that “Aneuploidy induction which does not involve the direct interaction of a chemical or its metabolite(s) with DNA is expected to have a threshold. This must be considered in the risk assessment for such chemicals: this is not addressed by current risk assessment guidelines”.

Parry et al. [7] proposed that chemicals exhibiting aneuploidy probably have thresholds. They suggested that experiments need to be conducted over a range of concentrations where there was no aneu-

ploidy induction and then over concentrations over which there was dose-related induction of aneuploidy. Elhajouji et al. [8] argued for the existence of a threshold for chemically induced aneuploidy in human lymphocytes using the in vitro micronucleus test for a number of aneugens. Tinwall and Ashby [9] considered that they had demonstrated a threshold at dose levels below 0.01 mg/kg for vincristine in the in vivo micronucleus. Haynes et al. [10] were, however, unable to distinguish between various different statistical models including or excluding a threshold based upon their experimental studies.

3. Dose-response models, non-linearity and thresholds

Attempts to identify thresholds usually involve some experimental investigation of a dose or concentration–response relationship. Statistical models of a dose–response relationship are usually some form of:

$$Y = F(d, x_1, x_2, \dots, x_i) + E,$$

where Y is the biological response, F describes a function and d , the administered dose or concentration. x_1, x_2, \dots, x_i are covariates such as age, genotype and sex while E is the error term or background noise. Often, such relationships are constrained to be monotonic such that the response at any dose level is as high as or higher than that at the zero dose level. Such dose–response relationships can either include or exclude a threshold.

A linear equation is one where all the non-constant terms are first degree. For example: $Y = a + bx$. A non-linear equation is one where one or more of the non-constant terms are more than the first degree. For example: $Y = a + bx^2$. They may result in a curvi-linear relationship but this does not automatically imply a threshold. Non-linear relationships can either have a threshold or not. Fig. 1 illustrates two hypothetical non-linear relationships — one with a threshold, the other without. A threshold is defined as a concentration (X_t) below which no effect occurs.

In the non-linear example, this is expressed as:

$$Y = a \quad (\text{for } X = 0 \text{ to } X_t),$$

$$Y = a + b(X - X_t)^2 \quad (\text{for } X \geq X_t).$$

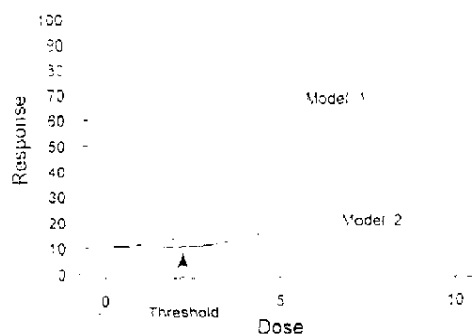


Fig. 1. Diagram to illustrate the difference between a non-linear curve without a threshold (Model 1) and a non-linear curve with a threshold (Model 2).

It should be noted that the term, non-linear, can be used in a number of different contexts. Elder and Kopp-Schneider [11], for instance, call a dose relationship with a threshold and a linear dose–response relationship (bX) above the threshold non-linear.

4. 'Pragmatic' vs. 'absolute' thresholds

A threshold model assumes that an effect only occurs at doses above a certain threshold dose level. ECETOC defined an 'absolute' threshold as "... a concentration below which a cell would not 'notice' the presence of the chemical. In other words, the chemical is present but does not interact with the cellular target". The precise identification of such a threshold, if it exists, is difficult [12]. A 'pragmatic' threshold can be considered as a concentration below which any effect is considered biologically unimportant (Fig. 2) [13]. This term is used in a somewhat similar way to how ECETOC defines a biological threshold except that it implies that there may be effects occurring because of treatment or exposure but these are considered below what might be considered biologically important. An example might be increases that did not exceed the range of responses seen in negative control material in a well-conducted series of experiments. Such a threshold may be defined, in part, with the help of statistical tests. The distinction between the various classifications of thresholds can initiate a philosophical discussion but is not relevant to regulatory risk assessment.

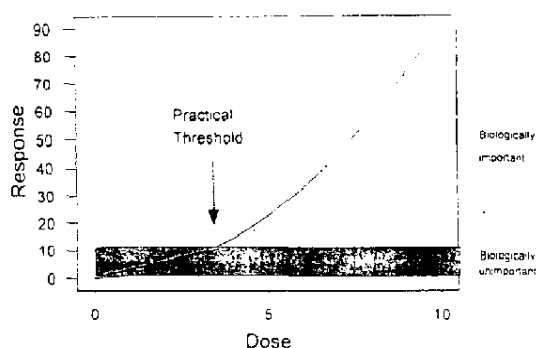


Fig. 2. Diagram to illustrate the concept of a biologically unimportant effect. The shaded area includes areas of the relationship where there is a dose-related increase but the response has not reached an effect considered of sufficient biological importance: for instance, has not exceeded the underlying background control incidence from a series of well-conducted experiments.

The use of safety factors (SFs) in toxicology to determine levels such as acceptable daily intakes (ADIs) (as opposed to the mathematically based low-dose linear extrapolation method) makes the assumption of a 'pragmatic' threshold based upon the identification of no or lowest observed (adverse) effect levels [NO(A)EL or LO(A)ELs] which are determined, to some extent, by statistical significance. The determination of NO(A)ELs from small experiments cannot, of course, guarantee zero risk at these levels of exposure. However, the experimental demonstration of these 'pragmatic' thresholds related to increased incidences considered to have no biological importance is easier than for an 'absolute' threshold for these non-stochastic effects [14].

5. Identification of thresholds by statistical tests: hypothesis testing vs. estimation

Statistical approaches to the identification of 'absolute' thresholds based upon the use of hypothesis testing are unsatisfactory. Conventionally, this approach involves comparing a null hypothesis such as a control = a treated group (i.e., a threshold exists) with an alternative hypothesis of a control < a treated group (i.e., no threshold exists). The 'poorer' the experiment, the harder it is to reject the null hypothesis and, as a consequence, to rule out the possibility

of ever smaller but real increases: the size of the experiment has to be increased disproportionately.

The problem is sometimes incorrectly described as the experimental designs used lacking sufficient statistical power to detect a threshold. However, the problem is, rather, that the null hypothesis of no difference in a comparison between two groups (i.e., below the threshold) can only be accepted provisionally. It cannot, therefore, be said [8] that the 'real biological threshold value for aneuploidy induction will basically be situated between the last statistically non-significant concentration and the first statistically highly significant concentration'. Such a statement might, however, provide an adequate description of the criteria for a 'pragmatic' threshold. A failure to detect an increase, defined by a statistically significant effect, is not proof of the absence of an effect at a low dosage. Determination of an increase, even if this does not reach statistical significance, may also mean that some biological events are actually happening at low doses which may be relevant to regulatory arguments relating to no-safe-doses and linearity at low dose. However, non-significant increases are often inappropriately referred to as negative results.

The definition of a 'pragmatic' threshold based upon a formal statistical test can also be misleading as the statistical significance is based upon the specific statistical test, experimental design and hypothesis being tested. A non-significant increase based upon a pairwise comparison might be significant using another test such as, for instance, a trend test. Statistical methods may, consequently, detect significant effects as a consequence of non-linearity but not rule out small but real increases at low doses.

Elhajouji et al. [8,15] provide examples of various modelling approaches for identifying thresholds for the induction of aneuploidy by a range of chemicals. Both papers provide numerous graphical examples of non-linear dose-response relationships which could be interpreted as evidence for a threshold. In the first paper, investigating *in vitro* micronuclei in human lymphocytes [8], a discontinuous regression model for the possible existence of a threshold was used which reflected a 'jump' in the regression line. A piecewise linear regression (or breakpoint) 'model' was fitted to the centromere-positive (MNCen+) vs. centromere-negative data. The breakpoint used for

the regression was the first statistically significant increase in MNCen+ (using Fisher's exact test to compare the micronuclei frequencies and the Chi-square test for centromere-positive frequencies between treated and control samples).

In the second paper [15] using fluorescence in situ hybridization (FISH) with human lymphocytes, they defined a threshold in two ways. Firstly, as the point where a statistically non-significant increase (the last statistically non-significant concentration, NOEL) changes to a statistically significant increase (the first statistically significant concentration, LOEL). A piecewise linear regression was carried out with the threshold defined as the first dose showing a significant increase. Secondly, by estimating inflection points (the point where there is a change from concavity to convexity of a curve) using a polynomial model as an approximate determination of a threshold.

Various curvi-linear models had previously been fitted to the data. The polynomial model was chosen as the best fitting model on the basis of goodness of fit or r^2 values and the least number of parameters needed. Mathematical models were also used to estimate the number of events ('hits') on the targets needed for the induction of chromosome loss and chromosome non-disjunction based upon multi-hit and multi-target equations [16]. It was noted that there were differences between estimates of the thresholds based upon the inflection point derived from the mathematical model and those estimated based upon the statistical tests. It is important to note that the experiments show a marked non-linearity of the dose-response relationship with a non-significant slight increase and then a highly significant steep increase. This seems good evidence for a 'pragmatic', but not necessarily an 'absolute', threshold.

An alternative approach to hypothesis testing is based upon the development of confidence intervals on the dose-response relationship. It may be possible, using this approach, to define the dose-response by estimating its size and shape. It would be possible to say, for instance, that evidence suggests that any increase, if it exists, is unlikely to exceed a certain value based upon, say, the upper 95% percentile confidence interval (CI). It may thus be possible to determine a 'pragmatic' threshold representing a non-linearity.

6. Dose-response models and curve fitting

The types of models available for fitting dose-response relationships can be broadly divided into four categories: empirical curve fits such as polynomials; semi-empirical models such as based upon the probit, logit or Weibull distributions; weak theoretical models such as the multi-hit or multi-stage models; and strong theoretical models such as the biologically-based dose-response (BB-DR) models. Weak and strong here referring to the quality of the underlying biological basis of the model.

A multitude of methods exist for empirical curve-fitting to data sets [17,18]. These include linear, log-linear, exponential, logit $\log(p/(1-p))$, multiple and logistical regression, polynomials, splines and non-parametric kernel density regression [19]. Confidence intervals can be determined by standard statistical procedures or procedures based upon re-sampling such as the Jackknife, Bootstrap and cross-validation. The STATISTICA software package, for instance, has a wide range of possible curve fits available while another package, SigmaPlot Version 3.0, claims to include 100 curve fits.

The majority of models used for curve fitting, however, have very little underlying biological basis. Some of the empirical models, such as the Weibull and Probit, are based upon an assumption of an underlying distribution of tolerances which has a superficial link to the concept of inter-individual variability in the thresholds for exposures to chemicals above which a toxicological event will result. The weaker theoretical models have the concepts of 'hits' or 'stages' which can be tenuously linked to biological models of damage to cellular targets or stages in tumour development. The statistical properties of many of these models have been recently reviewed [11].

The choice of curve-fitting model is often based upon fitting by 'eye' rather than by using any specific curve-fitting statistical program. Fitting a model by 'eye' can often guarantee a good fit to the data but many different 'eye' curves can often be fitted to the same data. It can then be difficult to distinguish between alternative models such as, for example, when a sigmoid curve (a non-threshold model) and a threshold-type model (Fig. 3) give equally good fits. As a result, it is not possible to discriminate between

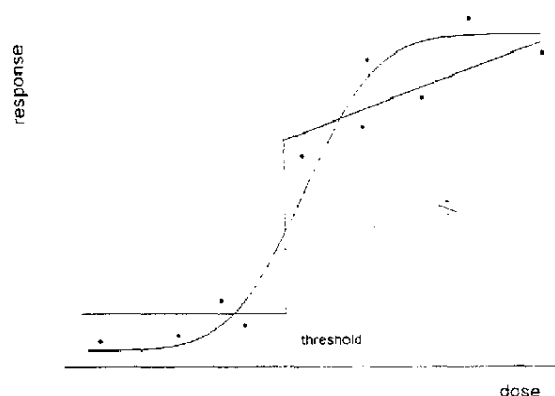


Fig. 3. A diagram showing an example where two different models — a sigmoid curve with no threshold and a model including a threshold — can be fitted by 'eye' to give similar good fits to the same data and showing the difficulty of distinguishing between alternative statistical models.

biological models based upon the statistical models. Consequently, the existence of a relationship chosen to show a good fit based upon the data does not prove the existence of a threshold or that the underlying mechanism postulated by the model reflects the actual mechanism.

Model fitting is usually assumed to be parsimonious in that the simplest model is chosen in preference to more complex ones: the principle of Occam's Razor — that the simplest explanation consistent with the known facts is chosen — is often evoked by modellers. It is sometimes argued [11] that a non-threshold model should be fitted first because it is the simpler model having less parameters; only if this model fails to fit the data should the threshold model then be considered. The choice of model is based upon some criteria of a goodness of fit test. These are often some measures such as a Chi-square or r^2 statistic. Post-hoc justification of a model — by determining through a statistical test that the fit is good — introduces a circular argument and does not prove that the model is necessarily correct.

7. Statistical models vs. biological models for thresholds: BB-DR models

The statistical models described previously are primarily data-driven, providing fits to the observed

data but with no or limited biological basis to the parameters in the model. The stronger theoretical BB-DR models have a more realistic basis, with the parameters in the model being mathematical representations of biological mechanisms. Example are the receptor-based models [20]. BB-DR models allow stochastic effects to be simulated and scenarios to be developed for extrapolation and prediction.

BB-DR offers the best prospect of providing realistic mathematical models of the underlying biological mechanisms. This could allow the development of hypotheses, for instance, of the action of a chemical on a biological target; the inclusion of non-linearity because of metabolic overload; thresholds because of the saturation of receptor occupancy, redundancy in the biological system or repair mechanisms.

BB-DR models are, however, data-intensive, requiring extensive experimental data to test their assumptions. The potential certainly exists for further development of these BB-DR models but the amount of work required to enable their use for identifying thresholds should not be underestimated.

8. Linearity at low dose and background incidence

The assumption of low-dose linearity underlies many mathematical modelling approaches to quantitative risk assessment of carcinogens. These methods make a number of assumptions. The dose-response relationship is assumed to be monotonic with the same mechanism of action resulting in both the background and the treated responses under consideration. Only a fraction of the background response has to be by this same mechanism for this generality to hold. If the same mechanism of action can be shown to be responsible for both control and treated responses, then it may be reasonable to assume no threshold. If the mechanism of action is not the same, then the assumption of no threshold is weaker. The assumption of low-dose linearity applies just as much to non-cancer as to cancer endpoints [21].

The linear-at-low-dose effect is a strong assumption. If the treatment incidence is raised or *in part* raised because of the same mechanism of action, then the argument about linear at low dose 'kicks

in'. In the context of thresholds for aneuploidy, it is relevant to note that the background level of spontaneous in vitro non-disjunction in the studies of Elhajjoui et al. [8] was 7%–15%. It may, thus, be difficult to argue for an 'absolute' threshold in these studies unless it can be shown that the induced non-disjunction occurs by a different mechanism in these studies.

The original proposal of identical mechanisms of action included when the concept of low-dose linearity was being proposed [22] may, however, now be too simple. With an increasing appreciation at the molecular level of the heterogeneity of the causes of toxicological events in both control and treated groups [23], it may be that the assumption only holds if the mechanisms are exactly the same, not just similar or variants of one another. Information on whether the underlying mechanisms responsible for spontaneous and induced aneuploidy are the same would thus provide some basis for deciding whether 'absolute' thresholds are possible for aneugens.

Most importantly and often understated are two further points: firstly, that the linearity only applies to very low doses (at the limit of the dose tending to zero) and appreciable non-linearity may apply at other dose levels; secondly, that the dose–response relationship at these low-dose levels may have a very low slope or potency and be very much lower (orders of magnitude) than that estimated based upon worst-case models which include data from high (or extremely high) dose levels.

The assumption of a linear model implies that there is no level of exposure to a 'high-dose' carcinogen that is not without some degree of risk at low dose. However, it is possible that low levels of exposure are, in fact, actually protective [24]. Hormesis implies the existence of a threshold below which effects are beneficial rather than harmful. This may result in what are termed J- or U- and β -shaped curves [13]. Some of the mathematical models used, such as the Linearized Multistage Model, constrain the parameters in the model to be ≥ 0 so that it is not easy to model, for instance, repair processes and hormesis. Removing this constraint would mean that if a repair process induced by low level exposure were to act in the same way as the background repair mechanism, then the model should predict protective effects which are also linear at low dose.

9. Experimental design for threshold studies

The philosophical difficulties in designing experiments to show 'absolute' thresholds were discussed earlier. However, the existence of sensitive and specific endpoints or biomarkers (for example, the use of selected chromosome probes in FISH) leading to low variability makes it possible to use a range of experimental designs for the characterization of any non-linearity in the dose–response relationship at low doses.

Two basic options exist: firstly, a hypothesis testing approach for whether there is a difference between the response in a negative control and a treatment; secondly, an estimation approach where the objective is to define and estimate the dose–response relationship. The experimental design — the number and spacing of doses and the allocation of experimental units between doses — depends upon the approach chosen and whether the data are qualitative or quantitative.

In vitro experiments may be able to detect quite small effects because of the sensitivity of some of the biomarkers available. However, careful experimental design with the use of the techniques of randomization and blocking is required to guard against artifactual results in the case of such sensitive endpoints. There is a need, in such studies, for both replication of individual units within experiments and repetition of experiments over time or place. In the case of in vivo studies, the inter-individual variability in response may be so large that any small but real effects at low dose will be swamped by the inherent biological variability of the system.

An important statistical issue is identifying what is the independent experimental unit in a study; whether, for instance, it is the culture or the cell in an in vitro experiment. For example, a study may consist of two independent cultures with 1000 cells scored per culture at each dose level. The choice of carrying out a statistical analysis based upon either 2000 cells or two cultures makes a large difference to the statistical power of the experiment. If, however, the cell is used as the experimental unit, as is often the case when Chi-square or Fisher's exact tests are used, and there are appreciable inter-culture differences, then the statistical significance of a hy-

pothesis test can be seriously wrong because assumptions underlying the test have been violated. Both the Chi-square and Fisher exact tests assume that all the experimental units, in this case the cells, are independent of one another. If there are culture differences, this assumption can be seriously violated.

The characterization of the dose–response relationships at low dose is an objective of the estimation approach. In vitro studies need to be, at least, repeated so that there is a measure of the inter-experimental variability in any dose–response relationship. It is important to identify the dose range where the threshold is likely to exist within or, alternatively, where the dose–response becomes highly non-linear. Dose levels should then span a range from zero (the negative control) to above the ‘pragmatic’ threshold. An important consideration is the use of extra replication around the area where the threshold is believed to occur and around doses where the accuracy of the dose–response relationship is important with the objective of improving the characterization of these regions. An appropriate biomarker should be chosen and, ideally, independent replicate measures obtained at each dose level. An appropriate dose–response model should be identified and confidence intervals, such as the 95% percentile, obtained for the relationship to provide some measure of the uncertainty associated with the relationship. The dose–response models could be a statistical model or, if sufficient previous work has been carried out, a BB-DR model. Power calculations should be carried out during the experimental design stage to help develop an optimal allocation of the resources in the experiment. The sensitivity of many of the possible biomarkers is such that experimental designs such as those based upon the factorial designs could be used to investigate more complex problems such as synergy and antagonism in the studies of mixtures.

10. Conclusions

Attempts are being made to harmonize different types of risk assessment with arguments made for the use of mathematical modelling approaches as opposed to the classic NO(A)EL/SF approach [25].

However, quantitative estimates of carcinogenic risks from extrapolation using a linear dose–response model have resulted in much more stringent regulation of carcinogens than non-carcinogens. Central to these mathematical approaches has been the underlying assumption of no threshold and linearity at low dose. The argument has been made that such assumptions are general and applicable to non-genotoxic endpoints just as much as genotoxic endpoints [21].

‘Absolute’ thresholds are, however, a philosophical problem in risk assessment. Statistical tests and mathematical models cannot prove or disprove their existence. Experiments to reject the hypothesis of linearity at low dose are impractical while a significance testing approach cannot be used to identify ‘absolute’ thresholds. It is possible, however, to define biologically unimportant increases and use this information to determine ‘pragmatic’ thresholds.

Estimation procedures can put the importance of thresholds, if any, into perspective. Experiments can be designed to estimate the dose–response relationships, providing measures of uncertainty rather than for formal hypothesis testing. Ideally, and increasingly in the future, such dose–response models should be based upon BB-DR modelling rather than empirical curve-fitting.

References

- [1] J.E. Coggle, *Biological Effects of Radiation*, 2nd ed., Taylor and Francis, London, 1983.
- [2] E. Efron, *The Apocalypses*, Simon and Schuster, New York, 1984.
- [3] C. Caulfield, *Multiple Exposures, Chronicles of the Radiation Age*, Penguin, Harmondsworth, Middlesex, 1990.
- [4] A.M. Weinberg, *Science and trans-science*, *Minerva* 10 (1972) 207–222.
- [5] V. Dellarco, K.H. Mavournim, R.R. Tice, Aneuploidy and health assessment: current status and future directions, *Environmental Mutagenesis* 7 (1985) 405–424.
- [6] M.J. Aardema, S. Albertini, P. Arno, L.M. Henderson, M. Kirsch-Volders, J.M. MacKay, A.M. Sarni, D.A. Stringer, R.D.F. Taalman, Aneuploidy: a report of an ECETOC task force, *Mutation Research* 410 (1998) 3–79.
- [7] J.M. Parry, R.J. Fielder, A. McDonald, Thresholds for aneuploidy-inducing chemicals, *Mutagenesis* 9 (1994) 503–504.
- [8] A. Elhajjoui, P. Van Hummelen, M. Kirsch-Volders, Indica-

- tions for a threshold of chemically induced aneuploidy in vitro in human lymphocytes, *Environmental and Molecular Mutagenesis* 26 (1995) 292–304.
- [9] H. Tinwall, J. Ashby, Micronucleus morphology as a means to distinguish aneugens and clastogens in the mouse bone marrow micronucleus assay, *Mutagenesis* 6 (1991) 193–198.
- [10] P. Haynes, T.R. Lambert, I. deG. Mitchell, Comparative in vivo genotoxicity of antiviral nucleoside analogues: penciclovir, acyclovir, ganciclovir and xanthine analogue, caffeine, in the mouse bone marrow micronucleus assay, *Mutation Research* 369 (1996) 65–74.
- [11] L. Elder, A. Kopp-Schneider, Statistical models for low-dose extrapolation, *Mutation Research* 405 (1998) 227–236.
- [12] M.A. Schneiderman, P. Decoufle, C.C. Brown, Thresholds for environmental cancer: biologic and statistical considerations, *Ann. N. Y. Acad. Sci.* 329 (1979) 92–130.
- [13] W.K. Lutz, Dose-response relationships in chemical carcinogenesis: superposition of different mechanisms of action, resulting in linear-non-linear curves, practical thresholds, J-shapes, *Mutation Research* 405 (1998) 117–124.
- [14] D.B. Clayson, D. Krewski, The concept of negativity in experimental carcinogenicity, *Mutation Research* 167 (1985) 233–240.
- [15] A. Elhajouji, F. Tibaldi, M. Kisch-Volders, Indications for thresholds of chromosome non-disjunction versus chromosome lagging induced by spindle inhibitors in vivo in human lymphocytes, *Mutagenesis* 12 (1997) 133–140.
- [16] A.M. Kellerer, Studies of the dose-effect relation, *Experientia* 45 (1989) 13–21.
- [17] P. Royston, D.G. Altman, Approximating statistical function by using fractional polynomial, *The Statistician* 46 (1997) 411–422.
- [18] E.J.T. Goetghebuer, S.J. Pocock, Detection and estimation of J-shaped risk-response relationships, *JRSS A* 185 (1995) 107–121.
- [19] B.W. Silverman, Some aspects of the spline smoothing approach to non-parametric regression curve fitting, *J. R. Stat. Soc. B* (1985) 1–52.
- [20] J.A. Gastel, T.R. Sutter, Biologically bounded risk assessment for receptor-mediated non-genotoxic carcinogens, *Regulatory Toxicology and Pharmacology* 22 (1995) 273–292.
- [21] M. Crawford, R. Wilson, Low-dose linearity: the rule or the exception, *Hum. Ecol. Risk Assess.* 2 (1996) 303–350.
- [22] K. Crump, D. Hoel, C. Langley, R. Peto, Fundamental carcinogenic processes and their implications to low-dose risk assessment, *Cancer Research* 36 (1976) 2973–2979.
- [23] K.W. Kinzler, B. Vogelstein, Gatekeepers and caretakers, *Nature* 386 (1997) 761–762.
- [24] K.T. Kitchen, J.L. Brown, Dose-response relationship for rat liver DNA damage caused by 49 rodent carcinogens, *Toxicology* 88 (1994) 31–49.
- [25] K.S. Crump, H.J. Clewell, M.E. Anderson, Cancer and non-cancer risk assessment should be harmonized, *Belle Newsletter* 5 (2/3) (1996) 2–4.